

## CHAPTER 4

### IMPLEMENTASI BIG DATA DAN KECERDASAN ARTIFISIAL UNTUK STATISTIK OFISIAL

**Imam Machdi<sup>1</sup>, Alfatihah Reno MNSPM<sup>2</sup>, Arie Wahyu Wijayanto<sup>3</sup>,  
Amanda Pratama Putra<sup>4</sup>, & Setia Pramana<sup>5</sup>**

Badan Pusat Statistik (BPS)

#### ABSTRAK

Badan Pusat Statistik (BPS) dalam memajukan *official statistics* menerapkan *corporate statistical infrastructure* yang mengadopsi tidak hanya sumber data administratif, sensus, dan survei, tetapi juga *big data*. Inovasi-inovasi pemanfaatan kecerdasan artifisial/*artificial intelligence* (AI) dan *big data* di BPS meliputi beberapa hal. Pertama, *mobile positioning data* (MPD) for *official statistics* yang memanfaatkan informasi lokasi dari perangkat *mobile* untuk pendukung sumber data statistik pariwisata serta deliniasi *metropolitan statistical area* di kota besar dan mengestimasi jumlah komuter. Kedua, pemanfaatan citra satelit untuk pertanian dan pemetaan kemiskinan. Pada statistik pertanian, pemanfaatan model AI digunakan untuk melakukan klasifikasi tutupan lahan pertanian dan estimasi luas tanam padi dan jagung. Sementara itu, pada pemetaan kemiskinan, model AI, seperti *convolutional neural network*, digunakan untuk mengekstraksi fitur-

---

<sup>1</sup> Imam Machdi menjabat sebagai Deputy Metodologi dan Informasi Statistik, Badan Pusat Statistik sejak November 2020 hingga sekarang. Gelar B.Sc diperoleh dari Louisiana State University, USA pada jurusan Computer Science. Beliau memperoleh gelar M.Sc dari *joint* program antara Institut Teknologi Sepuluh Nopember (ITS) dan Newcastle University, UK pada jurusan Computer Science. Kemudian pada tahun 2010, beliau mendapat gelar Ph.D in Engineering dari University of Tsukuba, Jepang di bidang Computer Science. Beliau bisa dikontak melalui e-mail machdi@bps.go.id

<sup>2</sup> Alfatihah Reno MNSPM, Pranata Komputer Ahli Madya pada Direktorat Sistem Informasi Statistik di Badan Pusat Statistik. Alfa memperoleh gelar Doktor di bidang Statistika di Institut Pertanian Bogor pada 2017. E-mail: alfa@bps.go.id

<sup>3</sup> Arie Wahyu Wijayanto saat ini merupakan dosen dan Kepala Pusat Penelitian dan Pengabdian kepada Masyarakat (PPPM), Politeknik Statistika STIS, Jakarta. Arie memperoleh gelar Doctor of Engineering bidang computer science dari Tokyo Institute of Technology pada 2020. Dapat dikontak pada e-mail: ariewahyu@stis.ac.id

<sup>4</sup> Amanda Pratama Putra, saat ini sebagai Statistisi Ahli Pertama pada Direktorat Pengembangan Metodologi Sensus dan Survei. Amanda menyelesaikan pendidikan D-IV di Sekolah Tinggi Ilmu Statistik, Jurusan Komputasi Statistik pada tahun 2016. E-mail: amanda.putra@bps.go.id

<sup>5</sup> Setia Pramana saat ini adalah Profesor bidang Statistika pada Politeknik Statistika STIS, Jakarta. Setia memperoleh gelar master dan Ph.D di bidang Bioinformatika dari Hasselt University, Belgium, dan dilanjutkan dengan program Postdoctoral di Karolinska Institutet, Stockholm Swedia. E-mail: setia.pramana@stis.ac.id

---

I. Machdi, A. R. MNSPM, A. W. Wijayanto, A. P. Putra, and S. Pramana  
Badan Pusat Statistik (BPS), e-mail: machdi@bps.go.id

@ 2023 Kolaborasi Riset dan Inovasi Industri Kecerdasan Artifisial (KORIKA) & Penerbit BRIN  
I. Machdi, A. R. MNSPM, A. W. Wijayanto, A. P. Putra, and S. Pramana, "Implementasi *big data* dan kecerdasan artifisial untuk statistik ofisial," in *Prosiding Use Cases Artificial Intelligence Indonesia: Embracing Collaboration for Research and Industrial Innovation in Artificial Intelligence*, B. R. Trilaksono, H. Riza, A. Jarin, N. D. S. Darmayanti, and S. Liawatimena, Eds. Jakarta: Penerbit BRIN, Februari 2023, ch. 4, pp. 47-53, doi: 10.55981/brin.668.c538  
ISBN: 978-623-8052-49-3, E-ISBN: 978-623-8052-50-9

fitur geospasial permukaan bumi, infrastruktur, dan pusat-pusat perekonomian dari citra siang dan malam hari untuk membantu memetakan sebaran kemiskinan di suatu wilayah. Selain itu, informasi dari berita *online* maupun media sosial digunakan untuk menangkap fenomena yang terjadi di masyarakat. Ke depannya, optimalisasi berbagai sumber data transaksi dan *big data*, dengan metodologi statistik dan *machine learning* akan ditingkatkan untuk menyempurnakan kualitas data yang dihasilkan BPS.

**Kata kunci:** big data, kecerdasan artifisial, statistik official

## A. PENDAHULUAN

Seiring cepatnya perkembangan digitalisasi di berbagai bidang, terjadi pertumbuhan data besar yang merupakan hasil interaksi miliaran orang yang menggunakan komputer, perangkat GPS, ponsel, perangkat medis, transaksi keuangan online atau seluler, dan media sosial.

Implementasi kecerdasan artifisial (AI) dan adopsi teknologi dan berbagai sumber *big data* merupakan tantangan sekaligus peluang bagi Badan Pusat Statistik (BPS) untuk meningkatkan kualitas statistik yang dihasilkan dengan memberikan data yang lebih cepat, lebih murah, dan lebih terperinci untuk melengkapi statistik official yang telah ada.

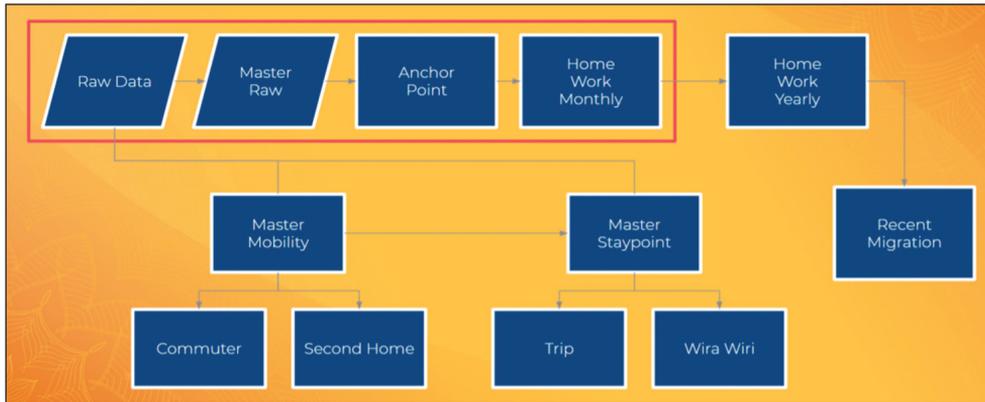
AI dapat diimplementasikan untuk menghasilkan statistik yang berasal dari data primer yang menggunakan model proses *generic statistical business process model* (GSBPM), dan juga untuk data sekunder (data administratif, *big data*, *non-probability sample*, dan sumber campuran). Berbagai sumber *big data* yang telah digunakan mulai dari *mobile positioning data* (MPD), citra satelit, media sosial, dan berita *online*.

Inisiatif pengembangan *big data* dan AI dilakukan oleh Direktorat Sistem Informasi Statistik, Politeknik Statistika STIS, Direktorat Analisis dan Pengembangan Model Statistik, serta berbagai divisi (*subject matter*) yang ada di BPS. Berikut ini *overview* singkat dari beberapa implementasi AI dan *big data* yang telah dilakukan.

## B. MOBILE POSITIONING DATA (MPD) UNTUK OFFICIAL STATISTICS

BPS telah menggunakan MPD sebagai salah satu sumber data sejak tahun 2016 yang diinisiasi untuk proyek penghitungan statistik wisatawan mancanegara di area perbatasan. Selanjutnya, pemanfaatan MPD terus diperluas untuk menghasilkan statistik wisatawan Nusantara, statistik perilaku komuter, statistik wisatawan mancanegara pada wilayah destinasi super prioritas (DSP), dan deliniasi wilayah metropolitan.

Analisis MPD dilakukan untuk mengembangkan *framework* algoritma yang disebut sebagai *anchor mobility data analytics* (AMDA). AMDA bertujuan mengubah *raw* data MPD menjadi *insight* untuk kemudian dirumuskan kembali sebagai



**Gambar 1.** *Framework* AMDA

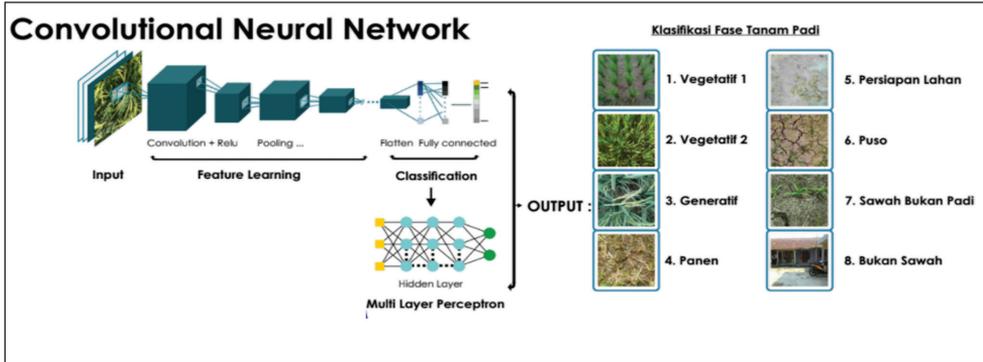
angka statistik sesuai dengan konsep definisi yang telah ditetapkan. Salah satu poin penting dalam AMDA adalah memetakan *usual environment* dari setiap pengguna berdasarkan data historis yang dimiliki. *Framework* AMDA dapat divisualisasikan komponen-komponennya pada Gambar 1.

Salah satu pergerakan manusia yang diamati melalui MPD adalah pola perilaku pariwisata. Konsep perjalanan (*trip*) yang digunakan adalah perjalanan yang dimulai dari lokasi *usual environment* menuju ke luar wilayah yang biasa dikunjungi hingga kembali lagi ke lokasi asal dengan berbasis pada batas wilayah administratif. Statistik yang dihasilkan berupa matriks kunjungan antarwilayah kabupaten/kota dan lama tinggal di daerah tujuan.

Pengamatan pergerakan manusia ini juga dimanfaatkan untuk menentukan deliniasi wilayah *metropolitan statistical area*. Pergerakan yang terjadi antara kecamatan dengan wilayah yang menjadi kota inti dihitung frekuensinya sebagai dasar penentuan kecamatan yang akan menjadi wilayah perluasan dari Kota Metropolitan. Tahun 2019 telah diselesaikan deliniasi Bandung Raya, dilanjutkan pada 2021 untuk wilayah Patungraya Agung dan Banjar Bakula.

### C. IMPLEMENTASI *DEEP LEARNING* UNTUK KLASIFIKASI FASE TANAM PADI

Dalam rangka membantu BPS pada survei kerangka sampel area (KSA) padi, dibangun aplikasi yang mengimplementasikan *deep learning* untuk mengklasifikasikan secara otomatis fase tanam padi ke dalam delapan kelas, yaitu vegetatif 1, vegetatif 2, generatif, panen, persiapan lahan, puso, sawah bukan padi, dan bukan sawah. Gambar 2 menunjukkan bagan proses klasifikasi ini yang menggunakan algoritma *convolutional neural network*. Akurasi ketepatan prediksi saat ini mencapai 74%.



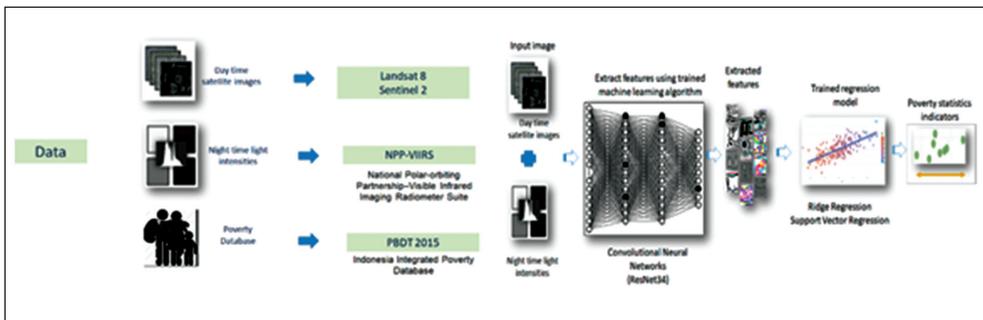
**Gambar 2.** Skema Klasifikasi Fase Tanam Padi dengan *Deep Learning*

#### D. PEMETAAN KEMISKINAN DENGAN CITRA SATELIT

Pemanfaatan *remote sensing* berdasarkan lokasi wilayah dengan atribut kondisi sosial ekonomi suatu wilayah melatarbelakangi pengembangan *remote sensing big data* untuk memprediksi lokasi kemiskinan.

Penggunaan *machine learning* serta lokasi citra yang menunjukkan aktivitas perekonomian suatu wilayah pada ukuran 1 km persegi dapat mendukung *official statistics* menyediakan data untuk kebijakan pengentasan kemiskinan dengan lebih cepat dan efisien.

Penggunaan intensitas cahaya malam sebagai pendekatan untuk mengukur pertumbuhan pendapatan daerah [1] mendasari pemodelan ini seperti ditampilkan pada Gambar 3. Intensitas cahaya malam dapat mewakili kondisi aktivitas sosial ekonomi [2] di suatu daerah, mengukur tingkat urbanisasi [3], pendapatan penduduk setempat [4], dan konsumsi listrik di suatu daerah [5][6].



**Gambar 3.** Data dan Algoritma Pemetaan Kemiskinan dengan Citra Satelit

Dalam memprediksi kemiskinan wilayah, digunakan *dataset* berupa citra satelit siang hari dan intensitas cahaya malam hari yang selanjutnya dilabelkan dengan tingkat kemiskinan dalam satuan ukuran *grid* 1 km persegi. Pada pembangunan model ini, digunakan data atribut kemiskinan yang tersedia pada tingkat rumah tangga melalui pendataan basis data terpadu (PBDT). Citra satelit siang yang digunakan berasal dari Satelit Landsat 8 dengan resolusi 30 m dan dan Sentinel 2 dengan resolusi 10 m. Sementara itu, data citra satelit malam menggunakan citra malam dari Satelit Suomi-NPP dengan menggunakan *visible infrared imaging radiometer suite* (VIIRS).

Proses ekstraksi citra satelit siang dan malam dalam bentuk *grid* sebelum dilakukan pemodelan dilakukan dengan *convolutional neural network* (CNN). Sementara itu, pada bagian pemodelan memprediksi lokasi kemiskinan digunakan *support vector machine* (SVM). SVM merupakan salah satu metode dalam *supervised learning* yang biasanya digunakan untuk klasifikasi (seperti *support vector classification*) dan regresi (*support vector regression*).

## **E. PREDIKSI INDEKS PEMBANGUNAN DESA (IPD) DENGAN MENGGUNAKAN CITRA SATELIT**

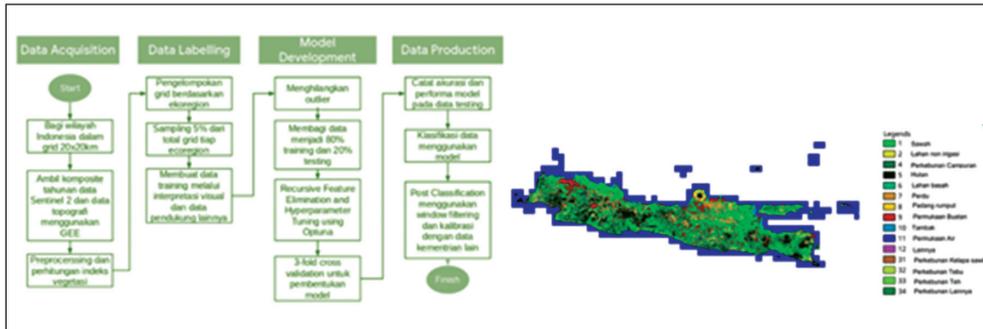
*Use case* ini mengombinasikan sumber *big data* citra satelit intensitas cahaya malam hari dengan teknik *transfer learning* untuk memprediksi status indeks pembangunan desa dengan mengasumsikan bahwa pembangunan infrastruktur dan aktivitas perekonomian malam hari menjadi prediktor nilai indeks pembangunan desa (IPD) [7]. Teknik *convolutional neural network* (CNN) dan *ridge regression* digunakan dalam metode penelitian ini.

## **F. KLASIFIKASI TUTUPAN LAHAN PERTANIAN UNTUK PERSIAPAN SENSUS PERTANIAN 2023**

Indonesia sebagai negara agraris memiliki berbagai tipe lahan pertanian. Citra satelit dapat merekam perubahan penggunaan lahan dari berbagai tipe lahan pertanian secara mutakhir dan mencakup seluruh wilayah di Indonesia.

Dengan menggunakan *machine learning* untuk mengklasifikasi data citra satelit ke dalam berbagai tipe lahan pertanian, BPS dapat menghasilkan kerangka statistik spasial berupa tutupan lahan pertanian untuk mendukung pelaksanaan sensus pertanian 2023.

Model klasifikasi tutupan lahan pertanian yang digunakan memiliki akurasi rata-rata lebih dari 93,6% (F-1 *score*). Sementara itu, akurasi di bawah 80% terjadi pada kelas tertentu, seperti kelas perkebunan campuran, karena tingginya heterogenitas karakteristik piksel.



**Gambar 4.** Tahapan Pengklasifikasian Tutupan Lahan Pertanian dan Contoh Visualisasinya

## G. MIND MAP GENERATOR OTOMATIS BERITA BENCANA ALAM INDONESIA

*Automatic mind map generator* dibangun untuk memahami isi suatu artikel kebencanaan dengan mengekstrak ide atau informasi utama dari sebuah artikel. Input dari alat ini adalah sebuah artikel dan output-nya adalah visualisasi *mind map*. Pembangunan *automatic mind map generator* membutuhkan model klasifikasi akurat yang digunakan untuk mengekstrak ide-ide utama dari artikel input. Model dapat mengklasifikasikan kalimat-kalimat dalam paragraf apakah itu mewakili informasi atau tidak. Artikel dapat menjadi informasi dengan struktur yang dapat menjelaskan *what, when, where, why, who, or how* suatu berita terkait bencana. Pemodelan artikel dilakukan menggunakan algoritma *SVM one vs rest* dengan kernel linier yang merupakan turunan dari SVM yang menyederhanakan klasifikasi *multiclass* menjadi klasifikasi biner.

## H. ANALISIS FENOMENA SOSIAL MEDIA DAN BERITA

Informasi tentang fenomena dapat dikumpulkan dari berbagai sumber, salah satunya berita dari portal berita *online* dan konten di media sosial. Informasi mengenai fenomena dari berita *online* dan media sosial dikumpulkan dan diolah menggunakan salah satu metode *artificial intelligence*, yakni *natural language processing* (NLP). Pengumpulan data dilakukan setiap triwulan dengan mengumpulkan berita dan konten *online* pada hasil pencarian media sosial dari 604 kata kunci yang telah ditentukan. Data berita *online* dikumpulkan oleh aplikasi Kofax Kapow, sedangkan data dari media sosial Twitter dikumpulkan dengan memanfaatkan fasilitas *application programming interface* (API) yang disediakan oleh Twitter.

Data dianalisis menggunakan metode *named entity recognition* (NER) dan metode analisis sentimen. Data fenomena yang telah dianalisis disajikan melalui platform *website* yang disebut “Fenomena BPS”. Di situs ini, pengguna (internal BPS) dapat mencari berita online dan konten media sosial serta fenomena yang ada di dalamnya.

## DAFTAR PUSTAKA

- [1] J. V. Henderson, A. Storeygard, dan D. N. Weil, "Measuring economic growth from outer space," *American Economic Review*, vol. 102, no. 2, pp. 994–1028, April 2012, doi: 10.1257/aer.102.2.994.
- [2] X. Chen dan W. D. Nordhaus, "Using luminosity data as a proxy for economic statistics," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 21, Mei 2011, pp. 8589–8594, doi: 10.1073/pnas.1017031108.
- [3] S. D. Miller, S. P. Mills, C. D. Elvidge, D. T. Lindsey, T. F. Lee, dan J. D. Hawkins, "Suomi satellite brings to light a unique frontier of nighttime environmental sensing capabilities," dalam *Proceedings of the National Academy of Sciences of the United States of America*, 2012 Sep 25, doi: 10.1073/pnas.1207034109
- [4] K. Ivan, I. H Holobăca, J. Benedek, dan I. Török, "VIIRS nighttime light data for income estimation at local level," *Remote Sensing*, vol. 12, no. 18, pp. 1–19, Sept. 2020, doi: 10.3390/rs12182950.
- [5] C. N. H. Doll dan S. Pachauri, "Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery," *Energy Policy*, vol. 38 no. 10, pp. 5661–5670, Okt. 2010, doi: 10.1016/j.enpol.2010.05.014.
- [6] C. D. Elvidge, K. E. Baugh, E. A Kihn, H. W. Kroehl, E. R. Davis, dan C. W. Davis, "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption," *International Journal of Remote Sensing*, Vol. 18, no. 6, pp. 1373–1379, Nov. 2010, doi: 10.1080/014311697218485.
- [7] D. Purnawanto, R. Nooraeni, dan N. W. K. Projo, "Estimating village development index based on satellite imagery using machine learning application," in *2020 International Conference on Applied Science and Technology (iCAST)*, 2020, pp. 527-531, doi: 10.1109/iCAST51016.2020.9557623.
- [8] D. Kristiyani dan A. W. Wijayanto, "Preserving women public restroom privacy using convolutional neural networks-based automatic gender detection," in *International Conference on Data Science and Official Statistics (ICDSOS) 2021*, 2021.
- [9] G. Tilottama, R. L. Powell, C. D. Elvidge, K. E. Baugh, P. C. Sutton, dan S. Anderson, "Shedding light on the global distribution of economic activity," *The Open Geography Journal*, vol. 3, no.1, pp. 1–31, Jan. 2010.
- [10] A. R. S Nugroho, A. R. M. N. S. P. Munaf, W. O. Z. Madjida, A. P. Putra, dan I. A. Setyadi, "Home and work identification process using mobile positioning data," in *Conference Of European Statisticians: Expert Meeting on Statistical Data Collection*, UNECE, Sept 2021.
- [11] P. D. Prabawa, H. T. Soblia, Y. F. Amin, W. Albertha, dan E. Seriwawan, "The use of mobile positioning data (MPD) to delineate metropolitan area in Indonesia: Case study in Cekungan Bandung" in *2020 Asia-Pacific Statistics Week*, Bangkok, Thailand, United Nations ESCAP, Juni 2020.